10/29/2020

MVA ASSIGNMENT 7

MEMBER INFORMATION

RUIXIN YANG (RUID: 197000459)

AISHWARYA SENTHILVEL (RUID: 199001269)

MULTIVARIATE REGRESSION ANALYSIS

Multivariate Regression is a method used to measure the degree at which more than one independent variable (predictors) and more than one dependent variable (responses), are linearly related.

```
> # Performing multiple regression on bank dataset
> fit2 = lm(y~Bank$age + pdays,data = bank)
> fit2

Call:
lm(formula = y ~ Bank$age + pdays, data = bank)

Coefficients:
(Intercept)     Bank$age        pdays
  0.0447758    0.0013896    0.0003333
```

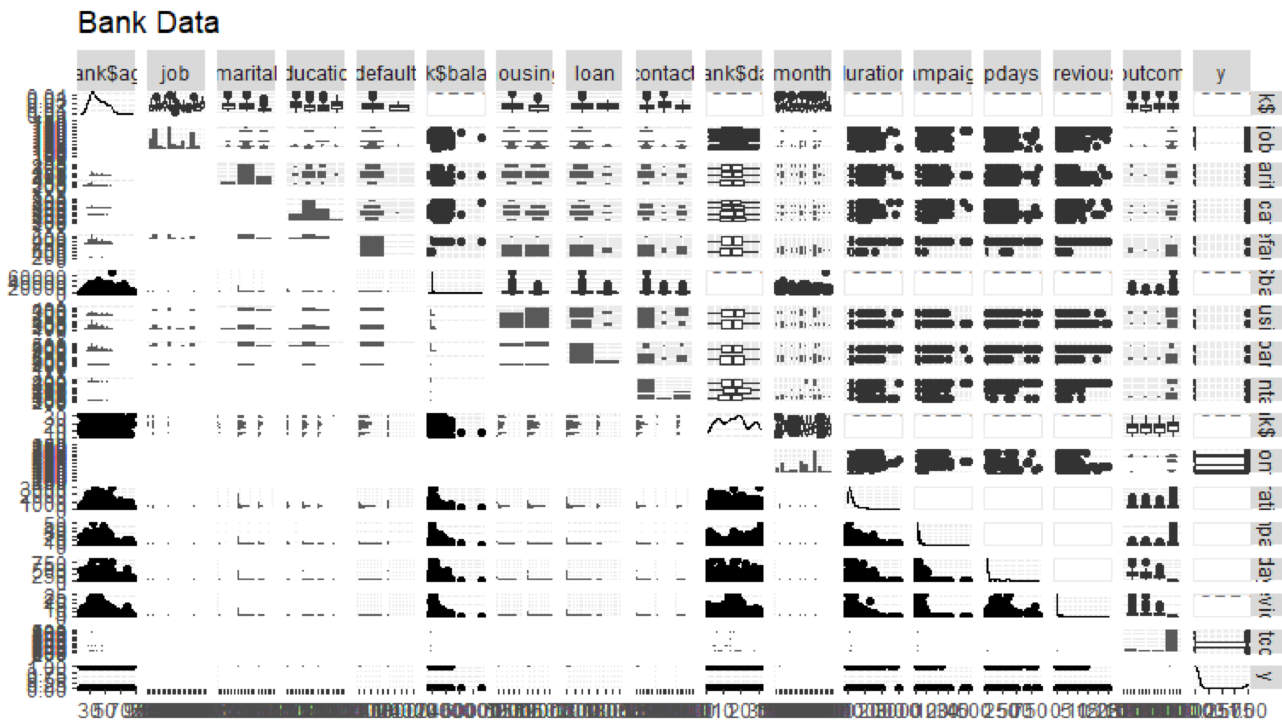
```
> coefficients(fit2)
 (Intercept)     Bank$age        pdays
0.0447757765 0.0013895996 0.0003333011
```

GGPAIRS

We are making a matrix of plots with the bank data set.

The ggpairs() function of the GGally package allows to build a great scatterplot matrix. Scatterplots of each pair of numeric variables are drawn on the left part of the figure.

Plot is shown below:



ANOVA FIT

Here we test if there is a difference between population means when a response variable is classified by one or more categorical variables (factors).

```
> #Anova Table
> anova(fit2)
Analysis of Variance Table

Response: y
            Df Sum Sq Mean Sq F value    Pr(>F)
Bank$age     1   0.94  0.9373  9.3069  0.002296 **
pdays        1   5.03  5.0330 49.9774 1.794e-12 ***
Residuals 4518 454.99  0.1007
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> confint(fit2,level=0.95)
                   2.5 %       97.5 %
(Intercept) 0.0073697805 0.0821817725
Bank$age    0.0005145957 0.0022646034
pdays       0.0002408708 0.0004257314
```

CONFINT

The “confint” methods calls the appropriate profile method, then finds the confidence intervals by interpolation in the profile traces

```
> confint(fit2,level=0.95)
                   2.5 %       97.5 %
(Intercept) 0.0073697805 0.0821817725
Bank$age    0.0005145957 0.0022646034
pdays       0.0002408708 0.0004257314
```

VCOV

Returns the variance-covariance matrix of the main parameters of a fitted model object.

```
> vcov(fit2)
              (Intercept)      Bank$age         pdays
(Intercept)  3.640437e-04 -8.208557e-06 -9.609780e-08
Bank$age    -8.208557e-06  1.992008e-07  1.871413e-10
pdays       -9.609780e-08  1.871413e-10  2.222797e-09
```

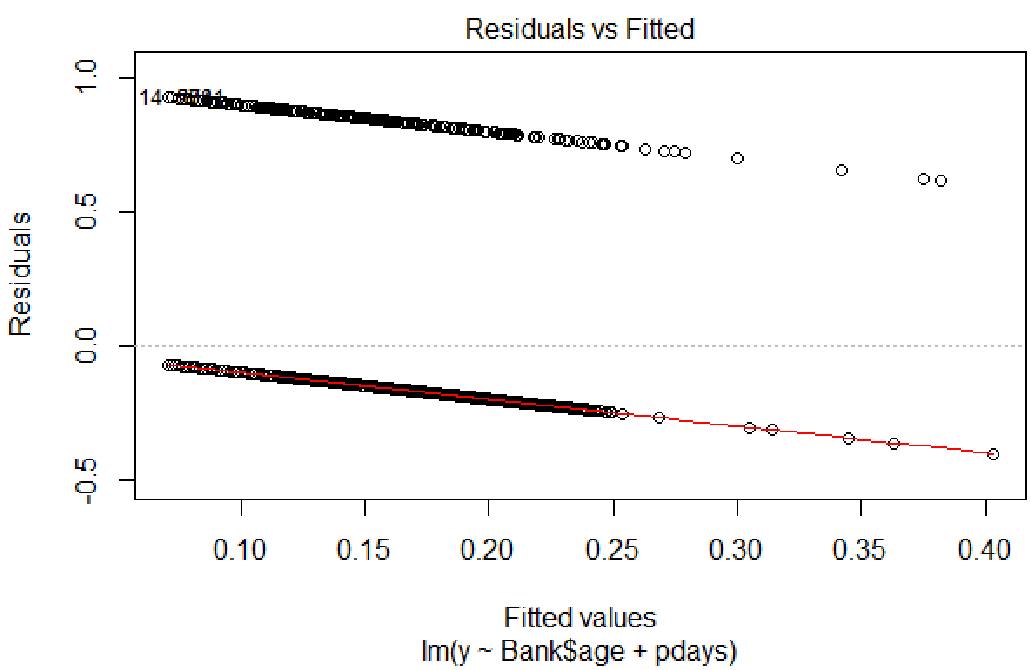
COV2COR

This converts a covariance matrix to a correlation matrix.
"Fit" is a symmetric numeric matrix, typically positive-definite since it often represents a covariance matrix.

```
> cov2cor(vcov(fit2))
             (Intercept)    Bank$age       pdays
(Intercept)   1.0000000 -0.96392786 -0.10682842
Bank$age     -0.9639279  1.00000000  0.00889353
pdays        -0.1068284  0.00889353  1.00000000
```

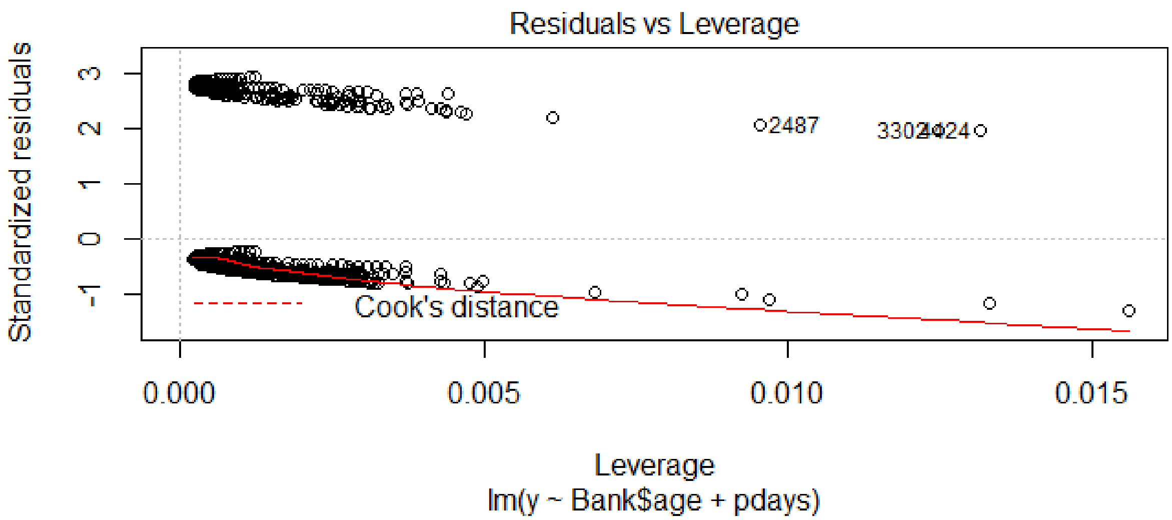
PLOTFIT

This function is useful for assessing how well a functional data object fits the actual discrete data.



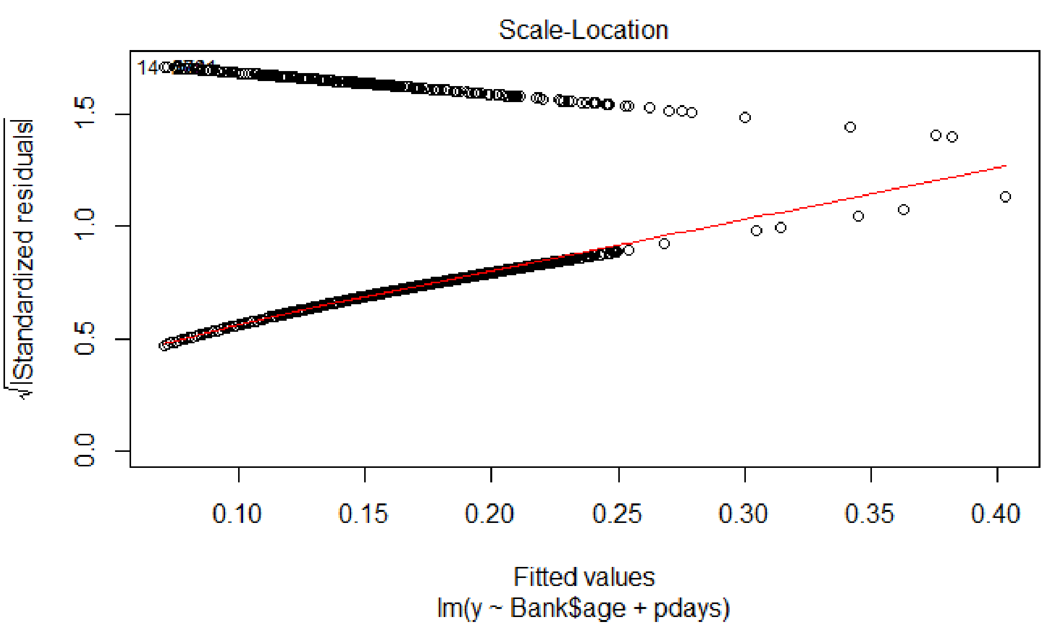
OUTLIER TEST

Here we assess outliers.



QQPLOT

Plots empirical quantiles of a variable, or of studentized residuals from a linear model, against theoretical quantiles of a comparison distribution.



COOK'S D PLOT

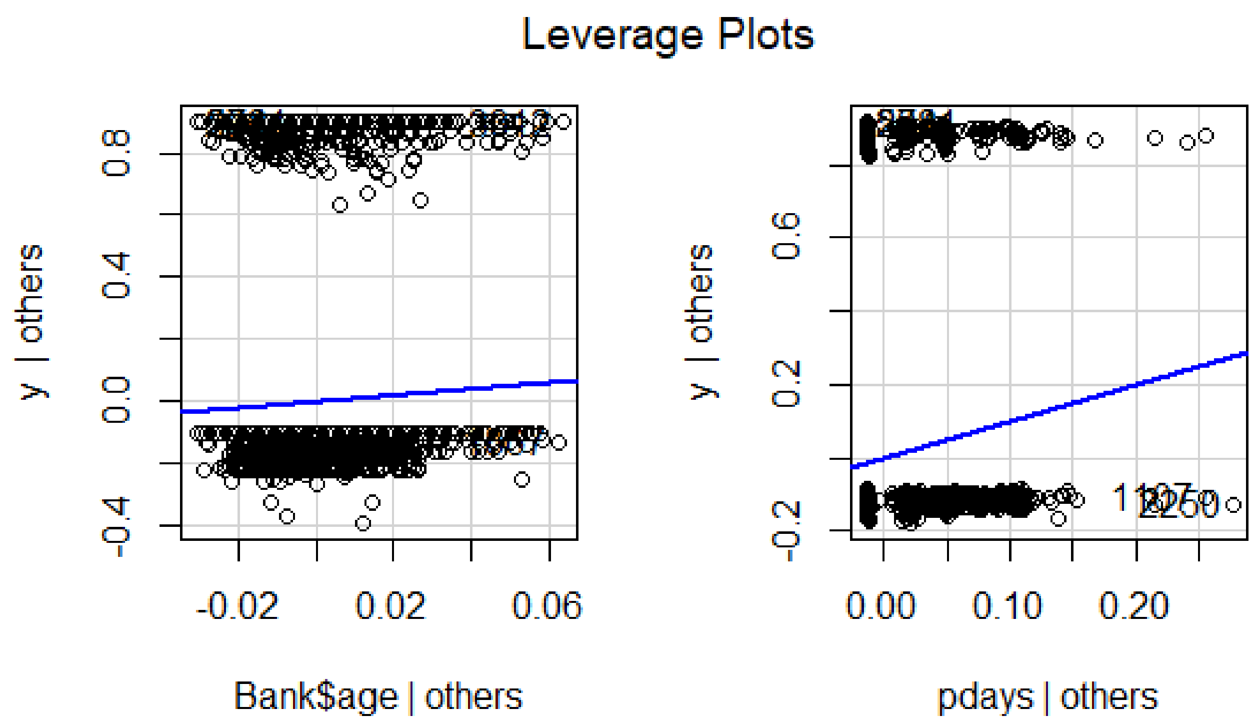
Cook's distance was introduced by American statistician R Dennis Cook in 1977. It is used to identify influential data points. It depends on both the residual and leverage. It considers both the x value and y value of the observation.

```
> # Cook's D plot
> # identify D values > 4/(n-k-1)
> cutoff <- 4/((nrow(bank)-length(fit2$coefficients)-2))
> plot(fit2, which=4, cook.levels=cutoff)
> cutoff
[1] 0.0008857396
```

LEVERAGE PLOTS

The leverage plot is a rescaled version of the usual added-variable plot.

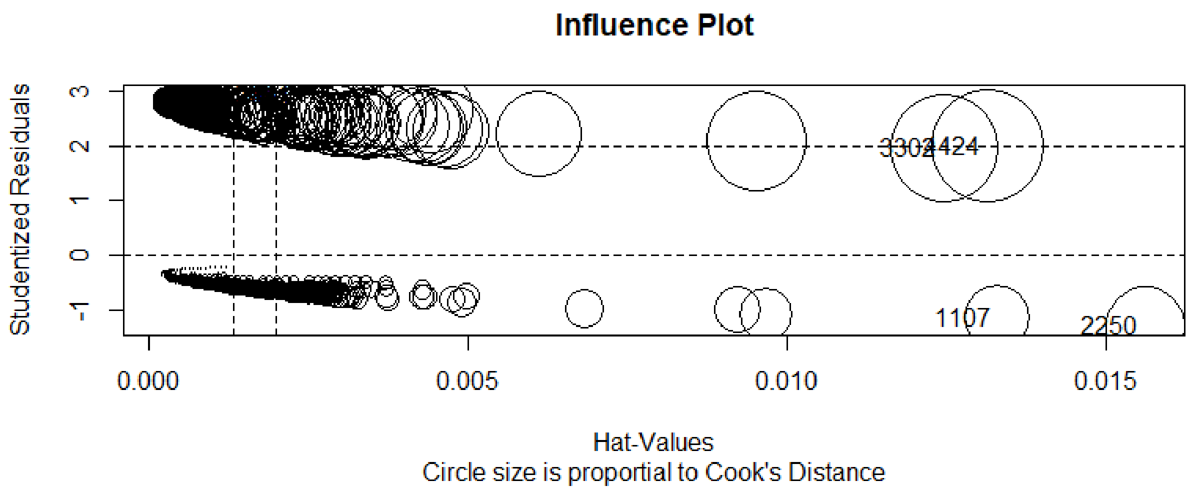
It is also called the partial-regression plot.



AVPLOTS

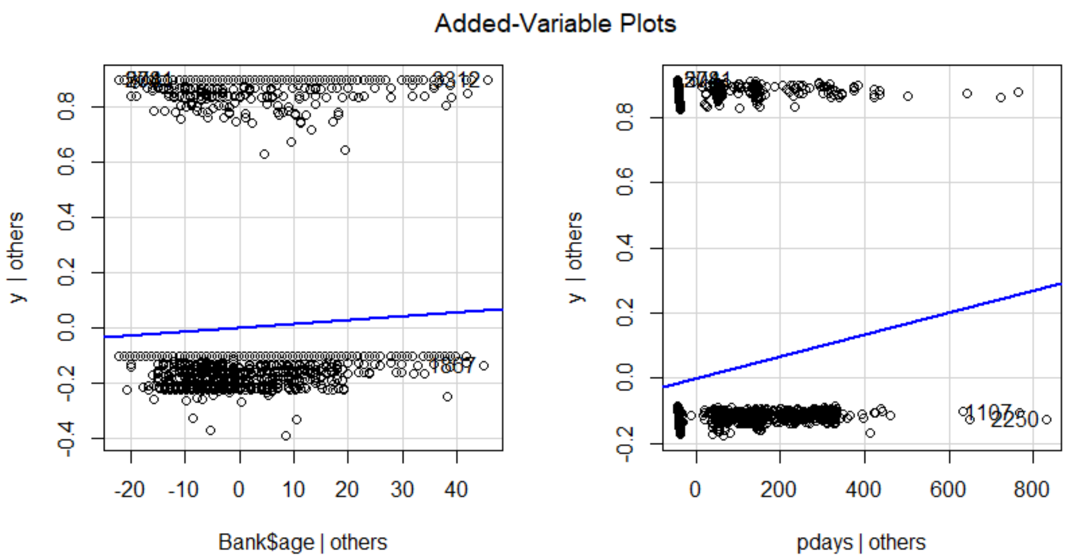
Creates an added-variable plot.

INFLUENCE PLOT




```
> # Influence Plot
> influencePlot(fit2, id.method="identify", main="Influence Plot", sub="Circle size is proportial to Cook's Distance")
        StudRes         Hat       CookD
504    2.932199 0.001233465 0.003533449
1107  -1.150713 0.013305469 0.005951535
2250  -1.280592 0.015617370 0.008671275
2781   2.932199 0.001233465 0.003533449
3302   1.960938 0.012454442 0.016154740
4424   1.982313 0.013152375 0.017445990
```

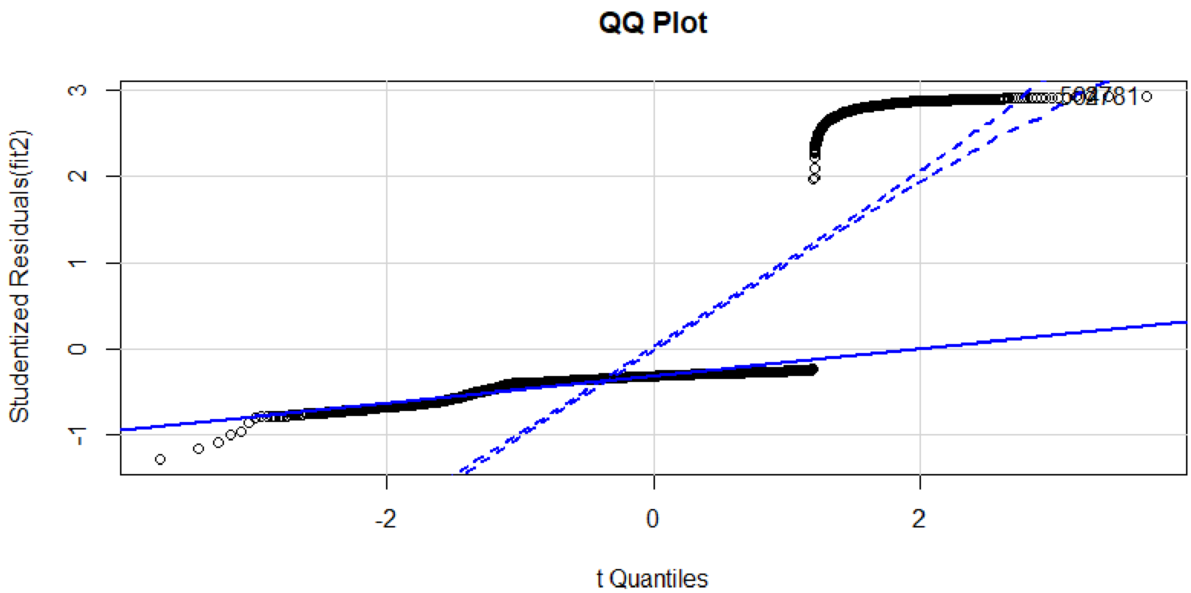
```
> cutoff <- 4/((nrow(bank)-length(fit2$coefficients)-2))
> cutoff
[1] 0.0008857396
```

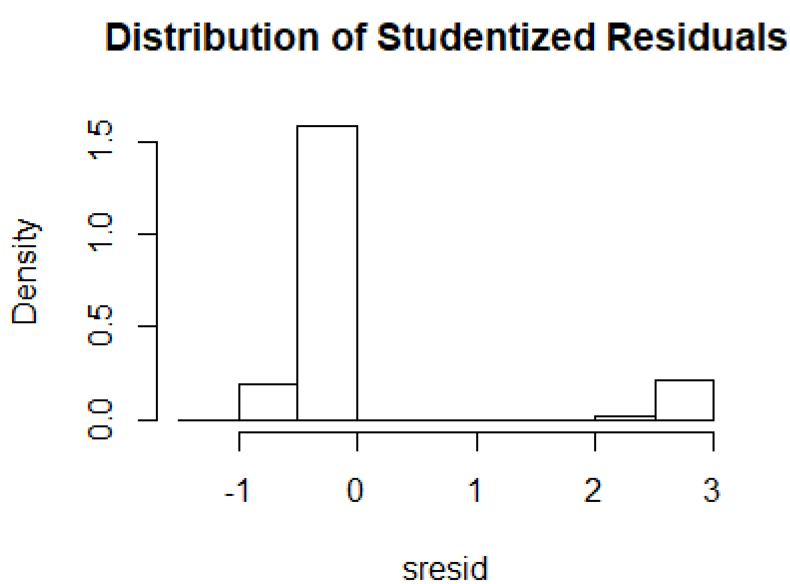


10/29/2020

QQPLOT

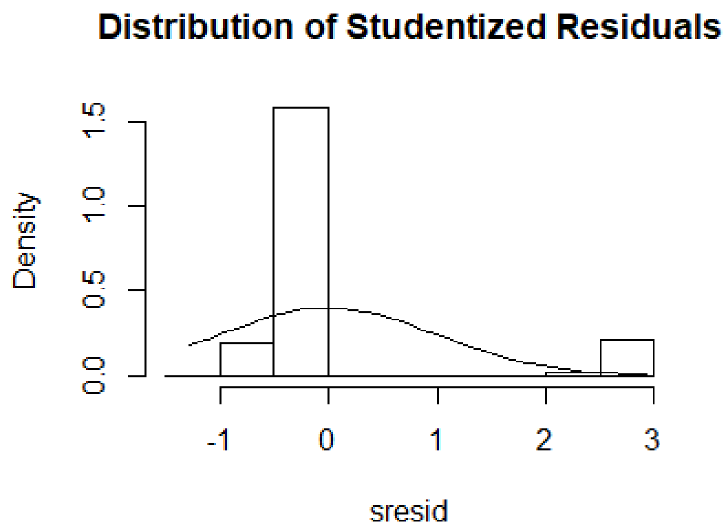
Normality of Residuals. QQ plot for studentized residuals.





10/29/2020

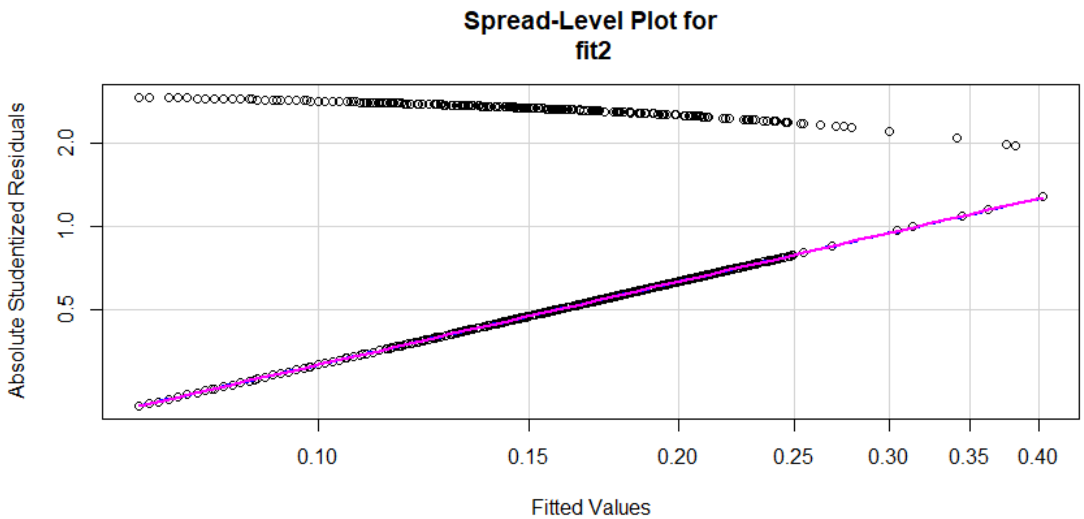
LINES XFIT-YFIT



VARIANCE INFLATION FACTORS

```
> # Evaluate Collinearity
> vif(fit2) # variance inflation factors
Bank$age    pdays
1.000079 1.000079
> sqrt(vif(fit2)) > 2 # problem?
Bank$age    pdays
   FALSE    FALSE
```

SPREAD LEVEL PLOT




```
> spreadLevelPlot(fit2)

Suggested power transformation:  -0.000988812
```

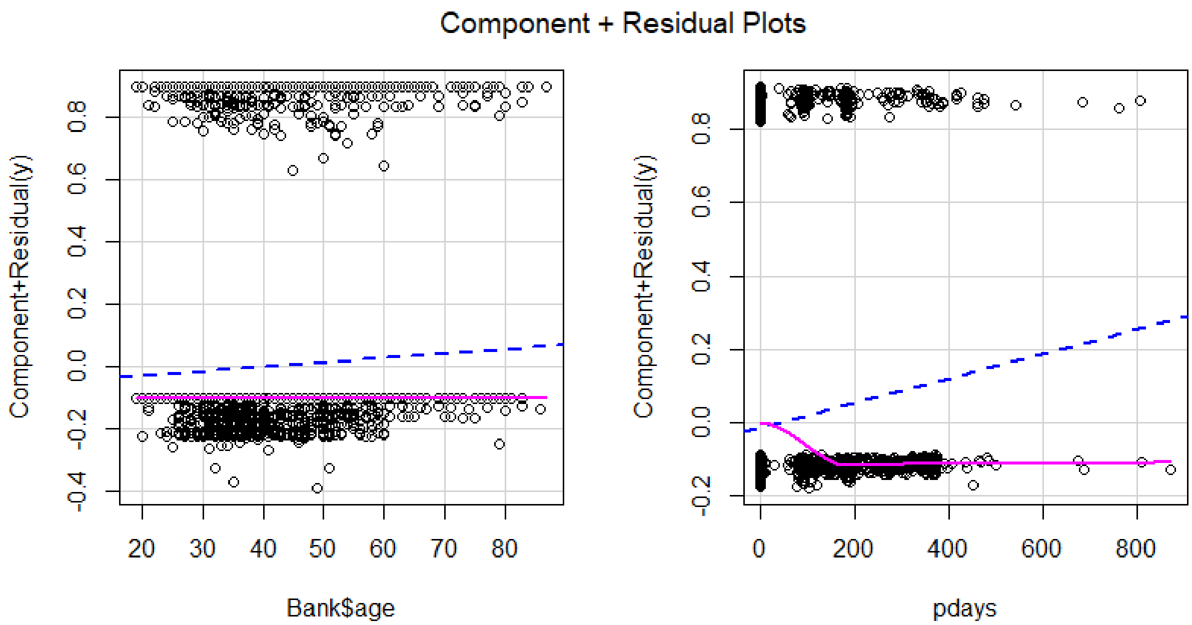
NCV TEST

Non-constant Variance Test

```
> # non-constant error variance test
> ncvTest(fit2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 142.3901, Df = 1, p = < 2.22e-16
```

DURBIN WATSON TEST

The Durbin–Watson statistic is a test statistic used to detect the presence of autocorrelation at lag 1 in the residuals from a regression analysis



LOGIT REGRESSION

After we replaced categorical regression by dummy variables, we ran a LOGIT regression to our new dataset since our dependent variable is binomial. Here is the result.

```
Call:
glm(formula = y ~ ., family = "binomial", data = bank)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0169  -0.3814  -0.2567  -0.1579   3.0346

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)        -2.462e+00  6.038e-01  -4.077 4.55e-05 ***
`Bank$age`         -4.232e-03  7.125e-03  -0.594 0.552537
`jobblue-collar`   -3.924e-01  2.420e-01  -1.621 0.104937
```

```
jobentrepreneur     -2.498e-01  3.811e-01  -0.655 0.512199    
jobhousemaid        -3.530e-01  4.176e-01  -0.845 0.398000    
jobmanagement       -7.302e-02  2.407e-01  -0.303 0.761602    
jobretired           6.315e-01  3.112e-01   2.029 0.042454 *  
`jobself-employed` -1.812e-01  3.533e-01  -0.513 0.608167    
jobservices         -1.457e-01  2.729e-01  -0.534 0.593542    
jobstudent           3.784e-01  3.750e-01   1.009 0.312958    
jobtechnician       -1.926e-01  2.301e-01  -0.837 0.402496    
jobunemployed       -6.395e-01  4.214e-01  -1.518 0.129138    
jobunknown           5.207e-01  5.853e-01   0.890 0.373669    
maritalmarried      -4.696e-01  1.743e-01  -2.694 0.007058 ** 
maritalsingle       -3.051e-01  2.038e-01  -1.497 0.134354    
educationsecondary   8.011e-02  2.022e-01   0.396 0.691924    
educationtertiary    3.208e-01  2.337e-01   1.373 0.169897    
educationunknown    -4.210e-01  3.572e-01  -1.179 0.238561    
default              5.446e-01  4.315e-01   1.262 0.206824    
`Bank$balance`      -3.911e-06  1.749e-05  -0.224 0.823014    
housing             -2.600e-01  1.381e-01  -1.883 0.059676 .  
loan                -6.296e-01  2.000e-01  -3.149 0.001640 ** 
contacttelephone    -7.020e-02  2.327e-01  -0.302 0.762900    
contactunknown      -1.416e+00  2.277e-01  -6.219 4.99e-10 ***
`Bank$day`           1.641e-02  8.161e-03   2.011 0.044362 *  
monthaug            -3.081e-01  2.494e-01  -1.235 0.216655    
monthdec             1.144e-01  6.573e-01   0.174 0.861784    
monthfeb             2.022e-01  2.937e-01   0.688 0.491290    
monthjan            -1.123e+00  3.816e-01  -2.944 0.003245 ** 
monthjul            -7.515e-01  2.498e-01  -3.008 0.002630 ** 
monthjun             5.542e-01  3.003e-01   1.845 0.065009 .  
monthmar             1.498e+00  3.901e-01   3.842 0.000122 ***
monthmay            -4.900e-01  2.340e-01  -2.094 0.036246 *  
monthnov            -8.430e-01  2.737e-01  -3.080 0.002072 ** 
monthoct             1.361e+00  3.300e-01   4.124 3.72e-05 ***
monthsep             6.572e-01  4.115e-01   1.597 0.110265    
duration             4.225e-03  2.020e-04  20.912  < 2e-16 ***
campaign            -7.042e-02  2.821e-02  -2.496 0.012549 *  
pdays               -9.791e-05  9.959e-04  -0.098 0.921684    
previous            -5.511e-03  3.818e-02  -0.144 0.885249    
poutcomeother        4.912e-01  2.692e-01   1.825 0.068019 .  
poutcomesuccess      2.445e+00  2.773e-01   8.818  < 2e-16 ***
poutcomeunknown     -1.216e-01  3.199e-01  -0.380 0.703822    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3231.0  on 4520  degrees of freedom
Residual deviance: 2173.7  on 4478  degrees of freedom
AIC: 2259.7

Number of Fisher Scoring iterations: 6
```

But the following code is not applicable to a logit regression. So, we change our estimation technique. We applied OLS estimation to the regression above. Here is the result.

```
Call:
lm(formula = y ~ ., data = bank)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.36460 -0.11175 -0.03981  0.02278  1.02817 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         6.594e-02  4.629e-02   1.424 0.154400    
`Bank$age`          1.478e-04  5.021e-04   0.294 0.768521    
`jobblue-collar`   -2.872e-02  1.604e-02  -1.791 0.073307 .  
jobentrepreneur    -1.929e-02  2.499e-02  -0.772 0.440204    
jobhousemaid       -3.096e-02  2.946e-02  -1.051 0.293316    
jobmanagement      -8.191e-03  1.745e-02  -0.469 0.638819    
jobretired          5.454e-02  2.421e-02   2.253 0.024293 *  
`jobself-employed` -1.039e-02  2.416e-02  -0.430 0.667059    
jobservices        -1.231e-02  1.823e-02  -0.675 0.499780    
jobstudent          4.913e-02  3.339e-02   1.471 0.141267    
jobtechnician      -1.874e-02  1.607e-02  -1.166 0.243549    
jobunemployed      -4.794e-02  2.736e-02  -1.752 0.079762 .  
jobunknown          4.190e-02  4.679e-02   0.896 0.370554    
maritalmarried     -3.941e-02  1.301e-02  -3.029 0.002465 ** 
maritalsingle      -2.214e-02  1.521e-02  -1.455 0.145618    
educationsecondary  6.847e-04  1.319e-02   0.052 0.958610    
educationtertiary   1.999e-02  1.615e-02   1.238 0.215710    
educationunknown   -2.993e-02  2.316e-02  -1.292 0.196433    
default             4.929e-02  3.157e-02   1.561 0.118580    
`Bank$balance`     -7.351e-07  1.376e-06  -0.534 0.593280    
housing            -1.837e-02  9.740e-03  -1.886 0.059400 .  
loan               -3.376e-02  1.152e-02  -2.931 0.003391 ** 
contacttelephone    4.874e-03  1.696e-02   0.287 0.773880    
contactunknown     -7.686e-02  1.384e-02  -5.555 2.93e-08 ***
`Bank$day`          1.546e-03  5.668e-04   2.728 0.006402 ** 
monthaug           -3.325e-02  2.061e-02  -1.613 0.106712    
monthdec            4.898e-02  6.326e-02   0.774 0.438776    
monthfeb            1.647e-02  2.531e-02   0.651 0.515263    
monthjan           -1.033e-01  2.821e-02  -3.662 0.000253 ***
monthjul           -6.469e-02  1.976e-02  -3.274 0.001070 ** 
monthjun            3.078e-02  2.350e-02   1.310 0.190310    
monthmar            2.200e-01  4.242e-02   5.187 2.23e-07 ***
monthmay           -3.416e-02  1.918e-02  -1.781 0.074944 .  
monthnov           -7.088e-02  2.136e-02  -3.319 0.000910 ***
monthoct            2.265e-01  3.476e-02   6.516 8.02e-11 ***
monthsep            9.746e-02  4.152e-02   2.347 0.018956 *  
duration            4.830e-04  1.564e-05  30.878  < 2e-16 ***
campaign           -1.430e-03  1.371e-03  -1.043 0.297112    
pdays              -5.226e-05  8.617e-05  -0.607 0.544187    
previous           -4.970e-04  3.276e-03  -0.152 0.879432    
poutcomeother       5.391e-02  2.307e-02   2.337 0.019473 *  
poutcomesuccess     4.197e-01  2.820e-02  14.884  < 2e-16 ***
poutcomeunknown    -2.416e-02  2.715e-02  -0.890 0.373535    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.2703 on 4478 degrees of freedom
Multiple R-squared:  0.2901,	Adjusted R-squared:  0.2834 
F-statistic: 43.56 on 42 and 4478 DF,  p-value: < 2.2e-16
```

Then we ran tests for assumptions of OLS estimation to see if they are satisfied. Here is the outcome.

```
Call:
lm(formula = y ~ ., data = bank)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.36460 -0.11175 -0.03981  0.02278  1.02817 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         6.594e-02  4.629e-02   1.424 0.154400    
`Bank$age`          1.478e-04  5.021e-04   0.294 0.768521    
`jobblue-collar`   -2.872e-02  1.604e-02  -1.791 0.073307 .  
jobentrepreneur    -1.929e-02  2.499e-02  -0.772 0.440204    
jobhousemaid       -3.096e-02  2.946e-02  -1.051 0.293316    
jobmanagement      -8.191e-03  1.745e-02  -0.469 0.638819    
jobretired          5.454e-02  2.421e-02   2.253 0.024293 *  
`jobself-employed` -1.039e-02  2.416e-02  -0.430 0.667059    
jobservices        -1.231e-02  1.823e-02  -0.675 0.499780    
jobstudent          4.913e-02  3.339e-02   1.471 0.141267    
jobtechnician      -1.874e-02  1.607e-02  -1.166 0.243549    
jobunemployed      -4.794e-02  2.736e-02  -1.752 0.079762 .  
jobunknown          4.190e-02  4.679e-02   0.896 0.370554    
maritalmarried     -3.941e-02  1.301e-02  -3.029 0.002465 ** 
maritalsingle      -2.214e-02  1.521e-02  -1.455 0.145618    
educationsecondary  6.847e-04  1.319e-02   0.052 0.958610    
educationtertiary   1.999e-02  1.615e-02   1.238 0.215710    
educationunknown   -2.993e-02  2.316e-02  -1.292 0.196433    
default             4.929e-02  3.157e-02   1.561 0.118580    
`Bank$balance`     -7.351e-07  1.376e-06  -0.534 0.593280    
housing            -1.837e-02  9.740e-03  -1.886 0.059400 .  
loan               -3.376e-02  1.152e-02  -2.931 0.003391 ** 
contacttelephone    4.874e-03  1.696e-02   0.287 0.773880    
contactunknown     -7.686e-02  1.384e-02  -5.555 2.93e-08 ***
`Bank$day`          1.546e-03  5.668e-04   2.728 0.006402 ** 
monthaug           -3.325e-02  2.061e-02  -1.613 0.106712    
monthdec            4.898e-02  6.326e-02   0.774 0.438776    
monthfeb            1.647e-02  2.531e-02   0.651 0.515263    
monthjan           -1.033e-01  2.821e-02  -3.662 0.000253 ***
monthjul           -6.469e-02  1.976e-02  -3.274 0.001070 ** 
monthjun            3.078e-02  2.350e-02   1.310 0.190310    
monthmar            2.200e-01  4.242e-02   5.187 2.23e-07 ***
monthmay           -3.416e-02  1.918e-02  -1.781 0.074944 .  
monthnov           -7.088e-02  2.136e-02  -3.319 0.000910 ***
monthoct            2.265e-01  3.476e-02   6.516 8.02e-11 ***
```

```
monthsep             9.746e-02  4.152e-02   2.347 0.018956 *  
duration             4.830e-04  1.564e-05  30.878  < 2e-16 ***
campaign            -1.430e-03  1.371e-03  -1.043 0.297112    
pdays               -5.226e-05  8.617e-05  -0.607 0.544187    
previous            -4.970e-04  3.276e-03  -0.152 0.879432    
poutcomeother        5.391e-02  2.307e-02   2.337 0.019473 *  
poutcomesuccess      4.197e-01  2.820e-02  14.884  < 2e-16 ***
poutcomeunknown     -2.416e-02  2.715e-02  -0.890 0.373535    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2703 on 4478 degrees of freedom
Multiple R-squared:  0.2901,    Adjusted R-squared:  0.2834 
F-statistic: 43.56 on 42 and 4478 DF,  p-value: < 2.2e-16


ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance =  0.05 

Call:
 gvlma(x = fit2) 

                      Value  p-value                   Decision
Global Stat        4218.368 0.000000 Assumptions NOT satisfied!
Skewness           1673.113 0.000000 Assumptions NOT satisfied!
Kurtosis           2532.447 0.000000 Assumptions NOT satisfied!
Link Function         9.579 0.001968 Assumptions NOT satisfied!
Heteroscedasticity    3.229 0.072347    Assumptions acceptable.
```

As you can see, the violation of assumptions for OLS estimation is severe. Only homoskedasticity holds for the regression. Then I compare these models above with analysis of variances. Here is the output.

```
       Df        Deviance          Resid. Df       Resid. Dev  
 Min.   :1   Min.   :  0.0241   Min.   :4478   Min.   :2174  
 1st Qu.:1   1st Qu.:  0.4844   1st Qu.:4488   1st Qu.:2962  
 Median :1   Median :  6.1400   Median :4499   Median :3097  
 Mean   :1   Mean   : 25.1750   Mean   :4499   Mean   :2961  
 3rd Qu.:1   3rd Qu.: 14.0654   3rd Qu.:4510   3rd Qu.:3169  
 Max.   :1   Max.   :598.2051   Max.   :4520   Max.   :3231  
 NA's   :1   NA's   :1                                       
```

After this, we applied the stepwise selection method for the former one with logit estimation. The chosen criterion is AIC. Then, we also applied the analysis of variances to this selected result. Here is the outcome.

```
                   Step Df    Deviance Resid. Df Resid. Dev      
 AIC
1                       NA          NA      4478   2173.651 225
9.651
2               - pdays  1 0.009673956      4479   2173.661 225
7.661
```

```
3           - previous  1 0.018483303      4480   2173.680 225
5.680
4           - monthdec  1 0.029477234      4481   2173.709 225
3.709
5     - `Bank$balance`  1 0.048284115      4482   2173.757 225
1.757
6    - contacttelephone  1 0.088933773      4483   2173.846 224
9.846
7       - jobmanagement  1 0.090235747      4484   2173.936 224
7.936
8  - `jobself-employed`  1 0.175514273      4485   2174.112 224
6.112
9        - jobservices  1 0.155507480      4486   2174.267 224
4.267
10 - educationsecondary  1 0.187242541      4487   2174.455 224
2.455
11    - poutcomeunknown  1 0.218682355      4488   2174.673 224
0.673
12    - jobentrepreneur  1 0.293672951      4489   2174.967 223
8.967
13      - jobtechnician  1 0.374951870      4490   2175.342 223
7.342
14          - monthfeb  1 0.377098038      4491   2175.719 223
5.719
15        - `Bank$age`  1 0.448453892      4492   2176.168 223
4.168
16      - jobhousemaid  1 0.625235658      4493   2176.793 223
2.793
17        - jobunknown  1 1.109219477      4494   2177.902 223
1.902
18           - default  1 1.381838320      4495   2179.284 223
1.284
19   - educationunknown  1 1.798365333      4496   2181.082 223
1.082
```

Then, we applied the subsets function “regsubsets” to our dataset. And here is the summary of output. It is also about variable selection. But its result is hard to interpret.

```
Subset selection object
Call: regsubsets.formula(y ~ ., data = bank, nbest = 10)
42 Variables  (and intercept)
                   Forced in Forced out
`Bank$age`              FALSE      FALSE
`jobblue-collar`        FALSE      FALSE
jobentrepreneur         FALSE      FALSE
jobhousemaid            FALSE      FALSE
jobmanagement           FALSE      FALSE
jobretired              FALSE      FALSE
`jobself-employed`      FALSE      FALSE
jobservices             FALSE      FALSE
jobstudent              FALSE      FALSE
jobtechnician           FALSE      FALSE
jobunemployed           FALSE      FALSE
jobunknown              FALSE      FALSE
```

```
maritalmarried          FALSE      FALSE
maritalsingle           FALSE      FALSE
educationsecondary      FALSE      FALSE
educationtertiary       FALSE      FALSE
educationunknown        FALSE      FALSE
default                 FALSE      FALSE
`Bank$balance`          FALSE      FALSE
housing                 FALSE      FALSE
loan                    FALSE      FALSE
contacttelephone        FALSE      FALSE
contactunknown          FALSE      FALSE
`Bank$day`              FALSE      FALSE
monthaug                FALSE      FALSE
monthdec                FALSE      FALSE
monthfeb                FALSE      FALSE
monthjan                FALSE      FALSE
monthjul                FALSE      FALSE
monthjun                FALSE      FALSE
monthmar                FALSE      FALSE
monthmay                FALSE      FALSE
monthnov                FALSE      FALSE
monthoct                FALSE      FALSE
monthsep                FALSE      FALSE
duration                FALSE      FALSE
campaign                FALSE      FALSE
pdays                   FALSE      FALSE
previous                FALSE      FALSE
poutcomeother           FALSE      FALSE
poutcomesuccess         FALSE      FALSE
poutcomeunknown         FALSE      FALSE
10 subsets of each size up to 8
Selection Algorithm: exhaustive
          `Bank$age` `jobblue-collar` jobentrepreneur jobhousema
id jobmanagement jobretired
1  ( 1 )  " "        " "              " "             " "
   " "           " "
1  ( 2 )  " "        " "              " "             " "
   " "           " "
1  ( 3 )  " "        " "              " "             " "
   " "           " "
1  ( 4 )  " "        " "              " "             " "
   " "           " "
1  ( 5 )  " "        " "              " "             " "
   " "           " "
1  ( 6 )  " "        " "              " "             " "
   " "           " "
1  ( 7 )  " "        " "              " "             " "
   " "           " "
1  ( 8 )  " "        " "              " "             " "
   " "           " "
1  ( 9 )  " "        " "              " "             " "
   " "           " "
1  ( 10 ) " "        " "              " "             " "
   " "           " "
```

```
2  ( 1 )  " "         " "               " "              " "
   " "          " "
2  ( 2 )  " "         " "               " "              " "
   " "          " "
2  ( 3 )  " "         " "               " "              " "
   " "          " "
2  ( 4 )  " "         " "               " "              " "
   " "          " "
2  ( 5 )  " "         " "               " "              " "
   " "          " "
2  ( 6 )  " "         " "               " "              " "
   " "          " "
2  ( 7 )  " "         " "               " "              " "
   " "          " "
2  ( 8 )  " "         " "               " "              " "
   " "          " "
2  ( 9 )  " "         " "               " "              " "
   " "          " "
2  ( 10 ) " "         " "               " "              " "
   " "          " "
3  ( 1 )  " "         " "               " "              " "
   " "          " "
3  ( 2 )  " "         " "               " "              " "
   " "          " "
3  ( 3 )  " "         " "               " "              " "
   " "          " "
          `jobself-employed` jobservices jobstudent jobtechnicia
n jobunemployed jobunknown
1  ( 1 )  " "                " "         " "        " "
  " "          " "
1  ( 2 )  " "                " "         " "        " "
  " "          " "
1  ( 3 )  " "                " "         " "        " "
  " "          " "
1  ( 4 )  " "                " "         " "        " "
  " "          " "
1  ( 5 )  " "                " "         " "        " "
  " "          " "
1  ( 6 )  " "                " "         " "        " "
  " "          " "
1  ( 7 )  " "                " "         " "        " "
  " "          " "
1  ( 8 )  " "                " "         " "        " "
  " "          " "
1  ( 9 )  " "                " "         " "        " "
  " "          " "
1  ( 10 ) " "                " "         " "        " "
  " "          " "
2  ( 1 )  " "                " "         " "        " "
  " "          " "
2  ( 2 )  " "                " "         " "        " "
  " "          " "
2  ( 3 )  " "                " "         " "        " "
  " "          " "
```

```
2  ( 4 )  " "                 " "         " "        " "
  " "          " "
2  ( 5 )  " "                 " "         " "        " "
  " "          " "
2  ( 6 )  " "                 " "         " "        " "
  " "          " "
2  ( 7 )  " "                 " "         " "        " "
  " "          " "
2  ( 8 )  " "                 " "         " "        " "
  " "          " "
2  ( 9 )  " "                 " "         " "        " "
  " "          " "
2  ( 10 ) " "                 " "         " "        " "
  " "          " "
3  ( 1 )  " "                 " "         " "        " "
  " "          " "
3  ( 2 )  " "                 " "         " "        " "
  " "          " "
3  ( 3 )  " "                 " "         " "        " "
  " "          " "
          maritalmarried maritalsingle educationsecondary educat
iontertiary educationunknown default
1  ( 1 )  " "            " "           " "                " "
    " "              " "
1  ( 2 )  " "            " "           " "                " "
    " "              " "
1  ( 3 )  " "            " "           " "                " "
    " "              " "
1  ( 4 )  " "            " "           " "                " "
    " "              " "
1  ( 5 )  " "            " "           " "                " "
    " "              " "
1  ( 6 )  " "            " "           " "                " "
    " "              " "
1  ( 7 )  " "            " "           " "                " "
    " "              " "
1  ( 8 )  " "            " "           " "                " "
    " "              " "
1  ( 9 )  " "            " "           " "                " "
    " "              " "
1  ( 10 ) " "            " "           " "                " "
    " "              " "
2  ( 1 )  " "            " "           " "                " "
    " "              " "
2  ( 2 )  " "            " "           " "                " "
    " "              " "
2  ( 3 )  " "            " "           " "                " "
    " "              " "
2  ( 4 )  " "            " "           " "                " "
    " "              " "
2  ( 5 )  " "            " "           " "                " "
    " "              " "
2  ( 6 )  " "            " "           " "                " "
    " "              " "
```

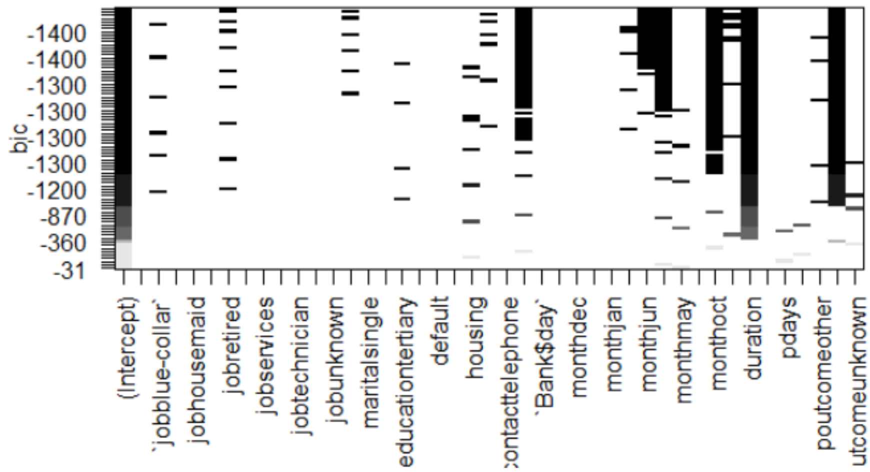
```
2  ( 7 )  " "           " "           " "                " "
            " "              " "
2  ( 8 )  " "           " "           " "                " "
            " "              " "
2  ( 9 )  " "           " "           " "                " "
            " "              " "
2  ( 10 ) " "           " "           " "                " "
            " "              " "
3  ( 1 )  " "           " "           " "                " "
            " "              " "
3  ( 2 )  " "           " "           " "                " "
            " "              " "
3  ( 3 )  " "           " "           " "                " "
            " "              " "
          `Bank$balance` housing loan contacttelephone contactun
known `Bank$day` monthaug monthdec
1  ( 1 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 2 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 3 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 4 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 5 )  " "           " "     " "  " "              "*"
      " "        " "      " "
1  ( 6 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 7 )  " "           "*"     " "  " "              " "
      " "        " "      " "
1  ( 8 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 9 )  " "           " "     " "  " "              " "
      " "        " "      " "
1  ( 10 ) " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 1 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 2 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 3 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 4 )  " "           " "     " "  " "              "*"
      " "        " "      " "
2  ( 5 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 6 )  " "           "*"     " "  " "              " "
      " "        " "      " "
2  ( 7 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 8 )  " "           " "     " "  " "              " "
      " "        " "      " "
2  ( 9 )  " "           " "     " "  " "              " "
      " "        " "      " "
```

```
2  ( 10 ) " "            " "      " "  " "               " "
     " "        " "       " "
3  ( 1 )  " "            " "      " "  " "               " "
     " "        " "       " "
3  ( 2 )  " "            " "      " "  " "               "*"
     " "        " "       " "
3  ( 3 )  " "            " "      " "  " "               " "
     " "        " "       " "
          monthfeb monthjan monthjul monthjun monthmar monthmay
monthnov monthoct monthsep duration
1  ( 1 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
1  ( 2 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 3 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 4 )  " "      " "      " "      " "      " "      " "
" "      "*"      " "      " "
1  ( 5 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 6 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 7 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 8 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      " "
1  ( 9 )  " "      " "      " "      " "      "*"      " "
" "      " "      " "      " "
1  ( 10 ) " "      " "      " "      " "      " "      "*"
" "      " "      " "      " "
2  ( 1 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 2 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 3 )  " "      " "      " "      " "      " "      " "
" "      "*"      " "      "*"
2  ( 4 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 5 )  " "      " "      " "      " "      "*"      " "
" "      " "      " "      "*"
2  ( 6 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 7 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 8 )  " "      " "      " "      " "      " "      "*"
" "      " "      " "      "*"
2  ( 9 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
2  ( 10 ) " "      " "      " "      " "      " "      " "
" "      " "      "*"      "*"
3  ( 1 )  " "      " "      " "      " "      " "      " "
" "      "*"      " "      "*"
3  ( 2 )  " "      " "      " "      " "      " "      " "
" "      " "      " "      "*"
```

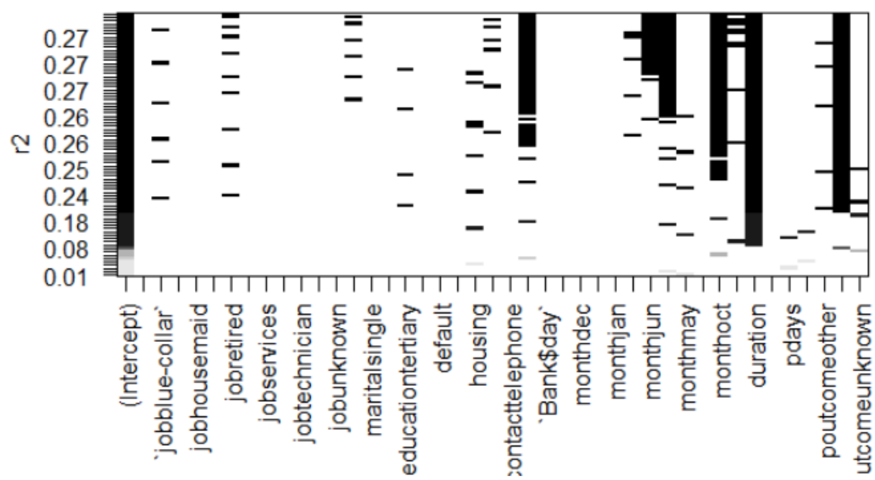
10/29/2020

```
3  ( 3 )  " "      " "      " "      " "      "*"      " "
" "      " "      " "      "*"
          campaign pdays previous poutcomeother poutcomesuccess
poutcomeunknown
1  ( 1 )  " "      " "   " "      " "           " "
" "
1  ( 2 )  " "      " "   " "      " "           "*"
" "
1  ( 3 )  " "      " "   " "      " "           " "
"*"
1  ( 4 )  " "      " "   " "      " "           " "
" "
1  ( 5 )  " "      " "   " "      " "           " "
" "
1  ( 6 )  " "      " "   "*"      " "           " "
" "
1  ( 7 )  " "      " "   " "      " "           " "
" "
1  ( 8 )  " "      "*"   " "      " "           " "
" "
1  ( 9 )  " "      " "   " "      " "           " "
" "
1  ( 10 ) " "      " "   " "      " "           " "
" "
2  ( 1 )  " "      " "   " "      " "           "*"
" "
2  ( 2 )  " "      " "   " "      " "           " "
"*"
2  ( 3 )  " "      " "   " "      " "           " "
" "
2  ( 4 )  " "      " "   " "      " "           " "
" "
2  ( 5 )  " "      " "   " "      " "           " "
" "
2  ( 6 )  " "      " "   " "      " "           " "
" "
2  ( 7 )  " "      " "   "*"      " "           " "
" "
2  ( 8 )  " "      " "   " "      " "           " "
" "
2  ( 9 )  " "      "*"   " "      " "           " "
" "
2  ( 10 ) " "      " "   " "      " "           " "
" "
3  ( 1 )  " "      " "   " "      " "           "*"
" "
3  ( 2 )  " "      " "   " "      " "           "*"
" "
3  ( 3 )  " "      " "   " "      " "           "*"
" "
 [ reached getOption("max.print") -- omitted 57 rows ]
```

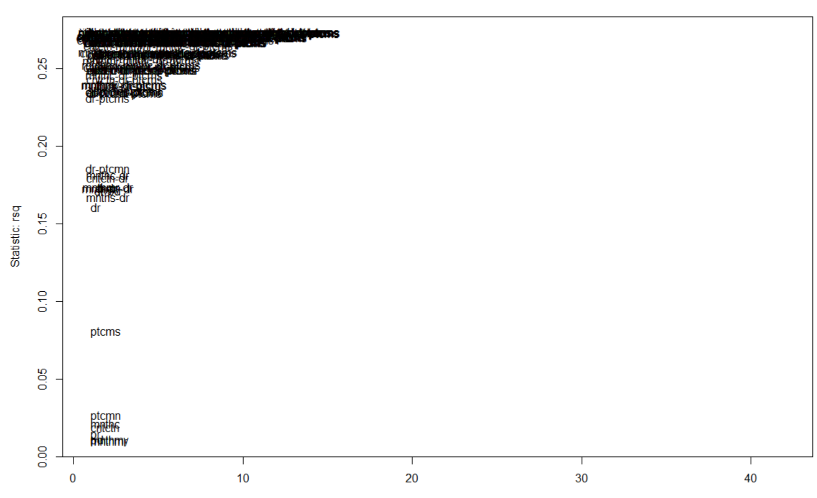
To interpret this result better, we generate a figure from the outcome above. Here is the result.



You can see the situation of choosing variables when it minimizes the BIC. We also replace BIC minimization by R2 maximization. Here is the result.



You can see that these results are nearly identical to each other. Then, we got the scatter plot of R2 when the subset size is growing. Here is the outcome.



You can see that there are some outliers when the size is relatively small. As the size grows, the R2 converges among different independent variables. Then we found ten pairs of coefficients when it ran regressions using these regressors separately. Here is the outcome.

```
[[1]]
  (Intercept)      duration 
-0.0148791849  0.0004929479 

[[2]]
    (Intercept) poutcomesuccess 
     0.09972678      0.54368408 

[[3]]
    (Intercept) poutcomeunknown 
      0.2254902      -0.1345320 

[[4]]
(Intercept)    monthoct 
  0.1089845   0.3535155 

[[5]]
   (Intercept) contactunknown 
    0.14388489    -0.09781238 

[[6]]
(Intercept)    previous 
 0.10329877  0.02200826 

[[7]]
(Intercept)     housing 
  0.1534149  -0.0674438 

[[8]]
 (Intercept)        pdays 
0.1020376378 0.0003319957 

[[9]]
(Intercept)    monthmar 
  0.1118068   0.3167646 

[[10]]
(Intercept)    monthmay 
 0.13704771 -0.07052411 
```

When we ran these regressions, we noticed that the ratio of 0 is larger than 0.5 in the dependent variable. That means most of clients did not purchase this deposit product. So, we chose the logit regression instead of the probit regression. But some of following functions are applicable to this model. Thus, we used the OLS estimation result for these functions. So, we calculated relative importance of these regressors. This function needs the normality assumption. It is violated in the logit regression. But we later found out that the burden of calculation is too much if we use the dataset. So, we drew a random sample by choosing the first 1000 clients' data. But it turns out that some matrices are singular in this case.

Finally, we changed the specification by dropping these dummies from categorical variables and preserving these numeric variables. That leaves us with six regressors. Here is the result of this regression.

```
Call:
lm(formula = bank$y ~ Bank$age + Bank$balance + Bank$day + Bank
$campaign +
    Bank$pdays + Bank$previous)

Residuals:
     Min       1Q   Median       3Q      Max
-0.52318 -0.11870 -0.09989 -0.08499  1.01514

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.273e-02  2.156e-02   2.445 0.014516 *
Bank$age       1.353e-03  4.465e-04   3.030 0.002462 **
Bank$balance   1.170e-06  1.569e-06   0.746 0.455995
Bank$day       3.014e-04  5.799e-04   0.520 0.603223
Bank$campaign -5.293e-03  1.538e-03  -3.442 0.000583 ***
Bank$pdays     1.662e-04  5.781e-05   2.876 0.004048 **
Bank$previous  1.573e-02  3.404e-03   4.622 3.91e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3163 on 4514 degrees of freedom
Multiple R-squared:  0.02045,   Adjusted R-squared:  0.01914
F-statistic:  15.7 on 6 and 4514 DF,  p-value: < 2.2e-16
```

The overall fitting situation is nearly acceptable. With this new specification, we calculated the relative importance of these regressors with four methods including LMG, last, first, and Pratt. Here is the result.

```
Response variable: bank$y
Total response variance: 0.1019823
Analysis based on 4521 observations

6 Regressors:
Bank$age Bank$balance Bank$day Bank$campaign Bank$pdays Bank$pre
vious
Proportion of variance explained by model: 2.04%
Metrics are normalized to sum to 100% (rela=TRUE).

Relative importance metrics:

                      lmg        last       first        pratt
Bank$age      0.098511128 0.178283016 0.066282933  0.098798221
Bank$balance  0.010226221 0.010795291 0.010451050  0.009655590
Bank$day      0.002554104 0.005248058 0.004121585 -0.004281359
Bank$campaign 0.146955525 0.230110483 0.121888562  0.154150734
Bank$pdays    0.301935795 0.160637700 0.353181540  0.265342431
Bank$previous 0.439817226 0.414925452 0.444074330  0.476334383

Average coefficients for different model sizes:
```

```
                         1X            2Xs            3Xs
  4Xs           5Xs           6Xs
Bank$age       1.361538e-03  1.359408e-03  1.358374e-03  1.35739
2e-03  1.355682e-03  1.352660e-03
Bank$balance   1.899874e-06  1.721354e-06  1.561243e-06  1.41694
2e-06  1.286799e-06  1.169917e-06
Bank$day      -4.353716e-04 -2.228967e-04 -4.730483e-05  9.57720
2e-05  2.106116e-04  3.014317e-04
Bank$campaign -6.279242e-03 -5.918759e-03 -5.648974e-03 -5.46070
3e-03 -5.344990e-03 -5.292993e-03
Bank$pdays     3.319957e-04  2.976939e-04  2.640154e-04  2.30910
7e-04  1.983351e-04  1.662465e-04
Bank$previous  2.200826e-02  2.065432e-02  1.935882e-02  1.81126
9e-02  1.690733e-02  1.573452e-02
```

Then we used bootstrap technique to get these measures of relative importance above with 1000 subsamples chose randomly. Here is the outcome.

```
Response variable: bank$y
Total response variance: 0.1019823
Analysis based on 4521 observations

6 Regressors:
Bank$age Bank$balance Bank$day Bank$campaign Bank$pdays Bank$pre
vious
Proportion of variance explained by model: 2.04%
Metrics are normalized to sum to 100% (rela=TRUE).

Relative importance metrics:

                      lmg        last       first         pratt
Bank$age      0.098511128 0.178283016 0.066282933  0.098798221
Bank$balance  0.010226221 0.010795291 0.010451050  0.009655590
Bank$day      0.002554104 0.005248058 0.004121585 -0.004281359
Bank$campaign 0.146955525 0.230110483 0.121888562  0.154150734
Bank$pdays    0.301935795 0.160637700 0.353181540  0.265342431
Bank$previous 0.439817226 0.414925452 0.444074330  0.476334383

Average coefficients for different model sizes:

                         1X            2Xs            3Xs
  4Xs           5Xs           6Xs
Bank$age       1.361538e-03  1.359408e-03  1.358374e-03  1.35739
2e-03  1.355682e-03  1.352660e-03
Bank$balance   1.899874e-06  1.721354e-06  1.561243e-06  1.41694
2e-06  1.286799e-06  1.169917e-06
Bank$day      -4.353716e-04 -2.228967e-04 -4.730483e-05  9.57720
2e-05  2.106116e-04  3.014317e-04
Bank$campaign -6.279242e-03 -5.918759e-03 -5.648974e-03 -5.46070
3e-03 -5.344990e-03 -5.292993e-03
Bank$pdays     3.319957e-04  2.976939e-04  2.640154e-04  2.30910
7e-04  1.983351e-04  1.662465e-04
```

```
Bank$previous  2.200826e-02  2.065432e-02  1.935882e-02  1.81126
9e-02  1.690733e-02  1.573452e-02


 Confidence interval information ( 1000 bootstrap replicates, bt
y= perc ):
Relative Contributions with confidence intervals:

                                       Lower  Upper
                    percentage 0.95   0.95    0.95
Bank$age.lmg         0.0985    _BCDEF  0.0056  0.2820
Bank$balance.lmg     0.0102    ___DEF  0.0003  0.0702
Bank$day.lmg         0.0026    ___DEF  0.0019  0.0566
Bank$campaign.lmg    0.1470    _BCD__  0.0532  0.2638
Bank$pdays.lmg       0.3019    ABC___  0.1440  0.4914
Bank$previous.lmg    0.4398    ABC___  0.2141  0.6201

Bank$age.last        0.1783    ABCDE_  0.0118  0.4295
Bank$balance.last    0.0108    ___DEF  0.0000  0.1076
Bank$day.last        0.0052    ___DEF  0.0000  0.1071
Bank$campaign.last   0.2301    ABCD__  0.0687  0.3962
Bank$pdays.last      0.1606    ABCDEF  0.0045  0.4986
Bank$previous.last   0.4149    ABCD__  0.0780  0.7068

Bank$age.first       0.0663    __CDEF  0.0040  0.2082
Bank$balance.first   0.0105    ___DEF  0.0000  0.0593
Bank$day.first       0.0041    ___DEF  0.0000  0.0546
Bank$campaign.first  0.1219    __CD__  0.0522  0.2143
Bank$pdays.first     0.3532    AB____  0.2205  0.4877
Bank$previous.first  0.4441    AB____  0.2776  0.5741

Bank$age.pratt       0.0988    _BCDE_  0.0055  0.2826
Bank$balance.pratt   0.0097    ___DEF -0.0007  0.0709
Bank$day.pratt      -0.0043    ___DEF -0.0054  0.0534
Bank$campaign.pratt  0.1542    _BCD__  0.0555  0.2742
Bank$pdays.pratt     0.2653    ABCD__  0.0364  0.5465
Bank$previous.pratt  0.4763    ABC___  0.1702  0.7383

Letters indicate the ranks covered by bootstrap CIs.
(Rank bootstrap confidence intervals always obtained by percenti
le method)
CAUTION: Bootstrap confidence intervals can be somewhat liberal.



 Differences between Relative Contributions:

                                                   Lower   Upper
                                  difference 0.95 0.95    0.95

Bank$age-Bank$balance.lmg           0.0883         -0.0216  0.267
3
Bank$age-Bank$day.lmg               0.0960         -0.0102  0.274
5
```

```
Bank$age-Bank$campaign.lmg        -0.0484          -0.2179  0.173
7
Bank$age-Bank$pdays.lmg           -0.2034          -0.4345  0.072
6
Bank$age-Bank$previous.lmg        -0.3413          -0.5747  0.019
1
Bank$balance-Bank$day.lmg          0.0077          -0.0463  0.061
6
Bank$balance-Bank$campaign.lmg    -0.1367     *    -0.2597 -0.027
9
Bank$balance-Bank$pdays.lmg       -0.2917     *    -0.4792 -0.115
9
Bank$balance-Bank$previous.lmg    -0.4296     *    -0.6122 -0.198
8
Bank$day-Bank$campaign.lmg        -0.1444     *    -0.2525 -0.034
3
Bank$day-Bank$pdays.lmg           -0.2994     *    -0.4866 -0.126
6
Bank$day-Bank$previous.lmg        -0.4373     *    -0.6069 -0.199
1
Bank$campaign-Bank$pdays.lmg      -0.1550          -0.3983  0.058
8
Bank$campaign-Bank$previous.lmg   -0.2929     *    -0.5321 -0.023
5
Bank$pdays-Bank$previous.lmg      -0.1379          -0.4433  0.229
6


Bank$age-Bank$balance.last         0.1675          -0.0285  0.411
5
Bank$age-Bank$day.last             0.1730          -0.0352  0.421
5
Bank$age-Bank$campaign.last       -0.0518          -0.3230  0.295
9
Bank$age-Bank$pdays.last           0.0176          -0.3902  0.327
2
Bank$age-Bank$previous.last       -0.2366          -0.6487  0.270
0
Bank$balance-Bank$day.last         0.0055          -0.0942  0.094
3
Bank$balance-Bank$campaign.last   -0.2193     *    -0.3817 -0.038
3
Bank$balance-Bank$pdays.last      -0.1498          -0.4868  0.029
5
Bank$balance-Bank$previous.last   -0.4041     *    -0.6942 -0.054
3
Bank$day-Bank$campaign.last       -0.2249     *    -0.3712 -0.038
2
Bank$day-Bank$pdays.last          -0.1554          -0.4842  0.036
2
Bank$day-Bank$previous.last       -0.4097     *    -0.6927 -0.052
8
Bank$campaign-Bank$pdays.last      0.0695          -0.3576  0.296
2
```

```
Bank$campaign-Bank$previous.last  -0.1848         -0.5775  0.183
2
Bank$pdays-Bank$previous.last     -0.2543         -0.6883  0.380
0


Bank$age-Bank$balance.first        0.0558         -0.0184  0.195
3
Bank$age-Bank$day.first            0.0622         -0.0143  0.199
3
Bank$age-Bank$campaign.first      -0.0556         -0.1774  0.108
9
Bank$age-Bank$pdays.first         -0.2869     *   -0.4504 -0.060
6
Bank$age-Bank$previous.first      -0.3778     *   -0.5436 -0.108
9
Bank$balance-Bank$day.first        0.0063         -0.0433  0.052
7
Bank$balance-Bank$campaign.first  -0.1114     *   -0.2099 -0.027
7
Bank$balance-Bank$pdays.first     -0.3427     *   -0.4811 -0.190
9
Bank$balance-Bank$previous.first  -0.4336     *   -0.5674 -0.261
0
Bank$day-Bank$campaign.first      -0.1178     *   -0.2093 -0.031
7
Bank$day-Bank$pdays.first         -0.3491     *   -0.4855 -0.198
1
Bank$day-Bank$previous.first      -0.4400     *   -0.5705 -0.261
2
Bank$campaign-Bank$pdays.first    -0.2313     *   -0.4026 -0.044
3
Bank$campaign-Bank$previous.first -0.3222     *   -0.4951 -0.109
9
Bank$pdays-Bank$previous.first    -0.0909         -0.3282  0.166
8


Bank$age-Bank$balance.pratt        0.0891         -0.0215  0.270
3
Bank$age-Bank$day.pratt            0.1031         -0.0066  0.279
6
Bank$age-Bank$campaign.pratt      -0.0554         -0.2277  0.171
2
Bank$age-Bank$pdays.pratt         -0.1665         -0.4706  0.129
8
Bank$age-Bank$previous.pratt      -0.3775         -0.6778  0.048
9
Bank$balance-Bank$day.pratt        0.0139         -0.0420  0.067
1
Bank$balance-Bank$campaign.pratt  -0.1445     *   -0.2698 -0.031
7
Bank$balance-Bank$pdays.pratt     -0.2557     *   -0.5398 -0.016
2
```

```
Bank$balance-Bank$previous.pratt  -0.4667     *   -0.7319 -0.143
6
Bank$day-Bank$campaign.pratt      -0.1584     *   -0.2727 -0.038
9
Bank$day-Bank$pdays.pratt         -0.2696     *   -0.5429 -0.022
2
Bank$day-Bank$previous.pratt      -0.4806     *   -0.7372 -0.165
3
Bank$campaign-Bank$pdays.pratt    -0.1112         -0.4415  0.150
9
Bank$campaign-Bank$previous.pratt -0.3222         -0.6448  0.011
1
Bank$pdays-Bank$previous.pratt    -0.2110         -0.6929  0.342
1

* indicates that CI for difference does not include 0.
CAUTION: Bootstrap confidence intervals can be somewhat liberal.
```

After this, we made predictions based on the revised regression above. And we display the first six predictions here since their variances are relatively small.

```
     1          2          3          4          5          6
0.09566587 0.22028637 0.17177573 0.07460014 0.12858233 0.1737543
8
```

Overall, these analyses are not so complete since we got several warnings during the process. And applying OLS estimation is somewhat inappropriate to a binary dependent variable. These flaws can be revised later.